

John Rawls (b. 1921, d. 2002) was an American political philosopher in the liberal tradition. His theory of *justice as fairness* describes a society of free citizens holding equal basic rights and cooperating within an egalitarian economic system. His theory of *political liberalism* explores the legitimate use of political power in a democracy, and envisions how civic unity might endure despite the diversity of worldviews that free institutions allow. His writings on *the law of peoples* set out a liberal foreign policy that aims to create a permanently peaceful and tolerant international order.

- 1. Life and Work
- 2. Aims and Method
  - 2.1 Four Roles of Political Philosophy
  - 2.2 The Sequence of Theories
  - 2.3 Ideal and Non-Ideal Theory
  - 2.4 Reflective Equilibrium
  - 2.5 The Independence of Moral and Political Theory
- 3. Political Liberalism: Legitimacy and Stability within a Liberal Society
  - 3.1 Legitimacy: The Liberal Principle of Legitimacy
  - 3.2 Reasonable Citizens
  - 3.3 Reasonable Pluralism and the Public Political Culture
  - 3.4 Political Conceptions of Justice
  - 3.5 Stability: An Overlapping Consensus
  - 3.6 Public Reason
- 4. Justice as Fairness: Justice within a Liberal Society
  - 4.1 The Basic Structure of Society
  - 4.2 Two Guiding Ideas of Justice as Fairness
  - 4.3 The Two Principles of Justice as Fairness
  - 4.4 The Conception of Citizens
  - 4.5 The Conception of Society
  - 4.6 The Original Position
  - 4.7 The Argument from the Original Position: The Selection of Principles
  - 4.8 The Argument from the Original Position: The Check for Stability
  - 4.9 Just Institutions: The Four-Stage Sequence
  - 4.10 The Original Position and Political Constructivism
- 5. The Law of Peoples: Liberal Foreign Policy
  - 5.1 The International Basic Structure and the Principles of the Law of Peoples
  - 5.2 Peoples: Liberal and Decent
  - 5.3 International Toleration and Human Rights
  - 5.4 The International Original Position
  - 5.5 Non-Ideal Theory: Outlaw States and Burdened Societies
  - 5.6 Reconciliation and Realistic Utopia
- 6. Further Reading
- Bibliography
  - Works by Rawls Cited in this Entry
  - Secondary Sources
- Academic Tools
- Other Internet Resources
- Related Entries

---

---

## 1. Life and Work

Rawls was born and raised in Baltimore, Maryland. His father was a prominent lawyer, his mother was a chapter president of the League of Women Voters. Rawls studied at Princeton and Cornell, where he was influenced by Wittgenstein's student Norman Malcolm; and at Oxford, where he worked with H. L. A. Hart, Isaiah Berlin, and Stuart Hampshire. His first professorial appointments were at Cornell and MIT. In 1962 Rawls joined the faculty at Harvard, where he taught for more than thirty years.

Rawls's adult life was a scholarly one: its major events occurred within his writings. The exceptions were two wars. As a college student, Rawls wrote an intensely religious senior thesis (*Bl*) and had considered studying for the priesthood. Yet Rawls lost his Christian faith as an infantryman in World War II on seeing the capriciousness of death in combat and learning of the horrors of the Holocaust. Then in the 1960s, Rawls spoke out against the draft for the Vietnam war because it discriminated against black and poor Americans. The Vietnam conflict impelled Rawls to analyze the defects in the American political system that led it to prosecute so ruthlessly what he saw as an unjust war, and to consider how citizens could conscientiously resist their government's aggressive policies.

Rawls's most discussed work is his theory of a just liberal society, called *justice as fairness*. Rawls first set out justice as fairness in systematic detail in his 1971 book, *A Theory of Justice*. Rawls continued to rework justice as fairness throughout his life, restating the theory in *Political Liberalism* (1993), *The Law of Peoples* (1999), and *Justice as Fairness* (2001).

This entry reflects Rawls's final statement of his views on justice as fairness, as well as on political liberalism and on the law of peoples. Some recent scholarship on Rawls's work can be found in [Further Reading](#) below.

## 2. Aims and Method

### 2.1 Four Roles of Political Philosophy

Rawls sees political philosophy as fulfilling at least four roles in a society's public culture. The first role is practical: philosophy can propose grounds for reasoned agreement when sharp political divisions threaten to lead to violent conflict. Rawls cites Hobbes's *Leviathan* as an attempt to solve the problem of order during the English civil war, Locke's *Letter on Toleration* as responding to the Wars of Religion, as well as the philosophy that emerged from the debates over the US Constitution, and from debates over the extension of slavery before the American civil war.

A second role of political philosophy is to help citizens to orient themselves within their own social world. Philosophy can meditate on what it is to be a member of a certain society—in a democracy, an equal citizen—and offer a unifying framework for answering divisive questions about how people with that political status should relate to each other.

A third role is to probe the limits of political possibility. Political philosophy must describe workable political arrangements that can gain support from real people. Yet within these limits, philosophy can be utopian: it can depict a social order that is the best that we can hope for. Given humans as they are, philosophy imagines laws as they might be.

A fourth role of political philosophy is reconciliation: "to calm our frustration and rage against our society and its history by showing us the way in which its institutions... are rational, and developed over time as they did to attain their present, rational form" (*JF*, 3). Philosophy can show that human life is not simply domination and cruelty, prejudice, folly and corruption; but that, at least in some ways, it is better that it has become as it is.

Rawls views his own work as a practical contribution to resolving the long-standing tension in democratic thought between liberty and equality, and to limning the limits of civic and of international toleration. He offers the members of democratic countries a way of understanding themselves as free and equal citizens of a society that is fair to all, and he describes a hopeful vision of a stably just constitutional democracy doing its part within a peaceful international community. To individuals who are frustrated that their fellow citizens and fellow humans do not see the whole truth as they do, Rawls offers the reconciling thought that this diversity of worldviews results from, and can support, a social order with greater freedom for all.

### 2.2 The Sequence of Theories

In contrast to the utilitarian, for Rawls political philosophy is not simply applied moral philosophy. The utilitarian holds to one universal moral principle ("maximize utility"), which she applies to individual actions, political constitutions, international relations, and all other subjects as required. Rawls has no universal principle: "The correct regulative principle for anything," he says, "depends on the nature of that thing" (*TJ*, 29). Rawls confines his theorizing to the political domain, and within this domain he holds that the correct principles for each sub-domain depend on its particular agents and constraints.

Rawls covers the domain of the political by addressing its sub-domains in sequence. The first sub-domain that he addresses is a self-contained democratic society reproducing itself across generations. Once principles are in place for such a society, Rawls moves to a second sub-domain: a society of nations, of which this democratic society is a member. Rawls suggests (though he does not show) that his sequence of theories could extend to cover further sub-domains, such as human

interactions with animals. Universal coverage will have been achieved once this sequence is complete, each sub-domain having received the principles appropriate to it.

## 2.3 Ideal and Non-Ideal Theory

Within each sub-domain of the political Rawls also follows a sequence: ideal theory before non-ideal theory. *Ideal theory* makes two types of idealizing assumptions about its subject matter. First, ideal theory assumes that all actors (citizens or societies) are generally willing to comply with whatever principles are chosen. Ideal theory thus idealizes away the possibility of law-breaking, either by individuals (crime) or societies (aggressive war). Second, ideal theory assumes reasonably favorable social conditions, wherein citizens and societies are able to abide by principles of political cooperation. Citizens are not so driven by hunger, for example, that their capacity for moral reasoning is overwhelmed; nor are nations struggling to overcome famine or the failure of their states.

Completing ideal theory first, Rawls says, yields a systematic understanding of how to reform our non-ideal world, and fixes a vision (mentioned above) of what is the best that can be hoped for. Once ideal theory is completed for a political sub-domain, *non-ideal theory* can be set out by reference to the ideal. For instance, once we find ideal principles for citizens who can be productive members of society over a complete life, we will be better able to frame non-ideal principles for providing health care to citizens with serious illnesses or disabilities. Similarly, once we understand the ideal principles of international relations, we will better see how the international community should act toward failed states, as well as toward aggressive states that threaten the peace.

## 2.4 Reflective Equilibrium

The aim of political philosophy is to reach justified conclusions about how political life should proceed. For Rawls, how justified one is in one's political convictions depends on how close one is to achieving *reflective equilibrium*. In reflective equilibrium all of one's beliefs, on all levels of generality, cohere perfectly with one another.

Thus, in reflective equilibrium one's specific political judgments (e.g., "religious intolerance is unjust," "racial discrimination is unjust") support one's more general political convictions (e.g., "all citizens have certain basic rights") which support one's very abstract beliefs about the political world (e.g., "all citizens are free and equal"). Viewed from the opposite direction, in reflective equilibrium one's abstract beliefs explain one's more general convictions, which in turn explain one's specific judgments. Were one to attain reflective equilibrium, the justification of each belief would follow from all beliefs relating in these networks of mutual support and explanation.

Though perfect reflective equilibrium is unattainable, one can use the *method of reflective equilibrium* to get closer to it and so increase the justifiability of one's beliefs. In carrying through this method, one begins with one's considered moral judgments: those made consistently and without hesitation when one is under good conditions for thinking (e.g., "slavery is wrong," "all citizens are political equals"). One treats these considered judgments as provisional fixed points, and then starts the process of bringing one's beliefs into relations of mutual support and explanation as described above. Doing this inevitably brings out conflicts where, for example, a specific judgment clashes with a more general conviction, or where an abstract principle cannot accommodate a particular kind of case. One proceeds by revising these beliefs as necessary, striving always to increase the coherence of the whole.

Carrying through this process of mutual adjustment brings one closer to *narrow reflective equilibrium*: coherence among one's initial beliefs. One then adds to this narrow equilibrium one's responses to the major theories in the history of political philosophy, as well as one's responses to theories critical of political philosophizing as such. One continues to adjust one's scheme of beliefs as one reflects on these alternatives, aiming for the end-point of *wide reflective equilibrium*, in which coherence is maintained after many alternatives have been considered.

Because of its emphasis on coherence, reflective equilibrium is often contrasted with foundationalism as an account of justified belief. Within foundationalist approaches, some subset of beliefs is considered to be unrevisable, thereby serving as a foundation on which all other beliefs are to be based. Reflective equilibrium privileges no such subset of beliefs: any belief at any level of generality is subject to revision, if revision will help to bring one's considered convictions into greater coherence overall.

## 2.5 The Independence of Moral and Political Theory

In working toward greater reflective equilibrium, any type of belief can in principle be relevant to one's conclusions about how political institutions should be arranged. Metaphysical beliefs about free will or personal identity might be relevant, as could epistemological beliefs about how we come to know what moral facts there are. However, while this is correct in principle, Rawls holds that in practice productive moral and political theorizing will proceed to a large extent independent of metaphysics and epistemology. Indeed, as a methodological presumption Rawls reverses the traditional order of priority. Progress in metaethics will derive from progress in substantive moral and political theorizing, instead of (as often assumed) vice versa (CF, 286–302).

Rawls's own metaethical theory of the objectivity and validity of political judgments, political constructivism, will be described below, after the substantive political theory from which it emerges.

### 3. Political Liberalism: Legitimacy and Stability within a Liberal Society

In a free society, citizens will have disparate worldviews. They will believe in different religions or none at all; they will have differing conceptions of right and wrong; they will disagree on how to live and on what relationships to value. Citizens will have contrary commitments, yet within any country there can only be one law. The law must either establish a national church, or not; women must either have equal rights, or not; abortion and gay marriage must either be permissible, or not; the economy must be set up in one way or another.

Rawls holds that the need to impose a unified law on a diverse citizenry raises two fundamental challenges. The first is the challenge of *legitimacy*: the legitimate use of coercive political power. How can it be legitimate to coerce all citizens to follow just one law, given that citizens will inevitably hold divergent worldviews?

The second challenge is the challenge of *stability*, which looks at political power from the receiving end. Why would a citizen willingly obey a law that is imposed on her by a collective body whose members have beliefs and values so different to her own? Yet unless most citizens willingly obey the law, no social order can be stable for long.

Rawls answers these challenges of legitimacy and stability with his theory of *political liberalism*. Political liberalism is not yet Rawls's theory of justice (justice as fairness). Political liberalism answers the conceptually prior questions of legitimacy and stability, so fixing the context and starting points for justice as fairness.

#### 3.1 Legitimacy: The Liberal Principle of Legitimacy

In a democracy, political power is always the power of the people as a collective body. In light of the diversity within a democracy, what would it mean for citizens legitimately to exercise coercive political power over one another? Rawls's test for the acceptable use of political power in a democracy is his *liberal principle of legitimacy*:

Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason (*PL*, 137).

According to this principle, political power may only be used in ways that all citizens can reasonably be expected to endorse. The use of political power must fulfill a *criterion of reciprocity*: citizens must reasonably believe that all citizens can reasonably accept the enforcement of a particular set of basic laws. Those coerced by law must be able to endorse the society's fundamental political arrangements freely, not because they are dominated or manipulated or kept uninformed.

The liberal principle of legitimacy intensifies the challenge of legitimacy: how can any particular set of basic laws legitimately be imposed upon a pluralistic citizenry? What constitution could all citizens reasonably be expected to endorse? Rawls's answer to this challenge begins by explaining what it means for citizens to be *reasonable*.

#### 3.2 Reasonable Citizens

Reasonable citizens want to live in a society in which they can cooperate with their fellow citizens on terms that are acceptable to all. They are willing to propose and abide by mutually acceptable rules, given the assurance that others will also do so. They will also honor these rules, even when this means sacrificing their own particular interests. Reasonable citizens want, in short, to belong to a society where political power is legitimately used.

Each reasonable citizen has her own view about God and life, right and wrong, good and bad. Each has, that is, what Rawls calls her own *comprehensive doctrine*. Yet because reasonable citizens are reasonable, they are unwilling to impose their own comprehensive doctrines on others who are also willing to search for mutually agreeable rules. Though each citizen may believe that she knows the truth about the best way to live, none is willing to force other reasonable citizens to live according to her beliefs, even if she belongs to a majority that has the power to enforce those beliefs on everyone. After all, Rawls says mentioning the Inquisition, oppressive use of state power will be necessary to

unite a society around any comprehensive doctrine, including the comprehensive liberalism of Kant or Mill (*PL*, 37).

One reason that reasonable citizens are so tolerant, Rawls says, is that they accept a certain explanation for the diversity of worldviews in their society. Reasonable citizens accept the *burdens of judgment*. The deepest questions of religion, philosophy, and morality are very difficult to think through. Even conscientious people will answer these questions in different ways, because of their particular life experiences (their upbringing, class, occupation, and so on). Reasonable citizens understand that these deep issues are ones on which people of good will can disagree, and so will be unwilling to impose their own worldviews on those who have reached conclusions different than their own.

### 3.3 Reasonable Pluralism and the Public Political Culture

Rawls's account of the reasonable citizen highlights his view of human nature. Humans are not irredeemably self-centered, dogmatic, or driven by what Hobbes called, "a perpetual and restless desire of power after power" (1651, 58). Humans have at least the capacity for genuine toleration and mutual respect.

This human capacity raises the hope that the diversity of worldviews in a democratic society may represent not merely pluralism, but *reasonable pluralism*. Rawls hopes, that is, that the religious, moral, and philosophical doctrines that citizens accept will themselves endorse toleration and accept the essentials of a democratic regime. In the religious sphere, for example, a reasonable pluralism might contain a reasonable Catholicism, a reasonable interpretation of Islam, a reasonable atheism, and so on. Being reasonable, none of these doctrines will advocate the use of coercive political power to impose religious conformity on citizens with different beliefs.

The possibility of reasonable pluralism softens but does not solve the challenge of legitimacy: how one law can legitimately be imposed on diverse citizens. For even in a society of reasonable pluralism, it would be unreasonable to expect everyone to endorse, say, a reasonable Catholicism as the basis for a constitutional settlement. Reasonable Muslims or atheists cannot be expected to endorse Catholicism as setting the basic terms for social life. Nor, of course, can Catholics be expected to accept Islam or atheism as the fundamental basis of law. *No* comprehensive doctrine can be accepted by all reasonable citizens, and so no comprehensive doctrine can serve as the basis for the legitimate use of coercive political power.

Yet where else then to turn to find the ideas that will define society's most basic laws, which all citizens will be required to obey? For Rawls, there is only one source of fundamental ideas that can serve as a focal point for all reasonable citizens of a liberal society. This is the society's *public political culture*.

Since justification is addressed to others, it proceeds from what is, or can be, held in common; and so we begin from shared fundamental ideas implicit in the public political culture in the hope of developing from them a political conception that can gain free and reasoned agreement in judgment (*PL*, 100-01).

The public political culture of a democratic society, Rawls says, "comprises the political institutions of a constitutional regime and the public traditions of their interpretation (including those of the judiciary), as well as historic texts and documents that are common knowledge" (*PL*, 13-14). Rawls looks to fundamental ideas implicit, for example, in the constitution's list of individual rights, in the design of the society's government, and in the historic decisions of important courts. These fundamental ideas from the public political culture can be crafted into a shared political conception of justice.

### 3.4 Political Conceptions of Justice

Rawls's solution to the challenge of legitimacy in a liberal society is for political power to be exercised in accordance with a *political conception of justice*. A political conception of justice is an interpretation of the fundamental ideas implicit in that society's public political culture.

A political conception is not derived from any particular comprehensive doctrine, nor is it a compromise among the worldviews that happen to exist in society at the moment. Rather, a political conception is freestanding: its content is set out independently of the comprehensive doctrines that citizens affirm. Reasonable citizens, who want to cooperate with one another on mutually acceptable terms, will see that a freestanding political conception generated from ideas in the public political culture is the only basis for cooperation that all citizens can reasonably be expected to endorse. The use of coercive political power guided by the principles of a political conception of justice will therefore be legitimate.

The three most fundamental ideas that Rawls finds in the public political culture of a democratic society are that citizens are *free* and *equal*, and that society should be a *fair* system of cooperation. All liberal political conceptions of justice will therefore be centered on interpretations of these three fundamental ideas.

Because there are many reasonable interpretations of “free,” “equal” and “fair,” there will be many liberal political conceptions of justice. Since all the members of this family interpret the same three fundamental ideas, however, all liberal political conceptions of justice will share certain basic features:

1. A liberal political conception of justice will ascribe to all citizens familiar individual rights and liberties, such as rights of free expression, liberty of conscience, and free choice of occupation;
2. A political conception will give special priority to these rights and liberties, especially over demands to further the general good (e.g., to increase national wealth) or perfectionist values (e.g., to promote a particular view of human flourishing);
3. A political conception will assure for all citizens sufficient all-purpose means to make effective use of their freedoms.

These abstract features must, Rawls says, be realized in certain kinds of institutions. He mentions several demands that all liberal conceptions of justice will make on institutions: a decent distribution of income and wealth; fair opportunities for all citizens, especially in education and training; government as the employer of last resort; basic health care for all citizens; and public financing of elections.

The use of political power in a liberal society will be legitimate if it is employed in accordance with the principles of any liberal conception of justice. By Rawls's criteria, a libertarian conception of justice (such as Nozick's in *Anarchy, State, and Utopia*) is not a liberal political conception of justice. Libertarianism does not assure all citizens sufficient means to make use of their basic liberties, and it permits excessive inequalities of wealth and power. By contrast, Rawls's own conception of justice (justice as fairness) does qualify as a member of the family of liberal political conceptions of justice.

## 3.5 Stability: An Overlapping Consensus

Political power is used legitimately in a liberal society when it is used in accordance with a political conception of justice. Yet the challenge of stability remains. Legitimacy means that the law may be enforced, but Rawls still needs to explain why citizens are willing to abide by it. If citizens do not believe that they have reasons to abide by the law from within their own perspectives, social order may disintegrate.

Rawls places his hopes for social stability on an *overlapping consensus*. In an overlapping consensus, citizens all endorse a core set of laws for different reasons. In Rawlsian terms, each citizen supports a political conception of justice for reasons internal to her own comprehensive doctrine.

Recall that the content of a political conception is freestanding: it is specified without reference to any comprehensive doctrine. This allows a political conception to be a “module” that can fit into any number of worldviews that citizens might have. In an overlapping consensus, each reasonable citizen affirms this common “module” from within her point of view.

Here is an example. The quotation below from the second Vatican Council of the Catholic Church shows how a particular comprehensive doctrine (Catholicism) affirms one component of a liberal political conception (a familiar individual liberty) for its own reasons:

This Vatican Council declares that the human person has a right to religious freedom. This freedom means that all men are to be immune from coercion on the part of individuals or of social groups and of any human power, in such wise that in matters religious no one is forced to act in a manner contrary to his own beliefs. Nor is anyone to be restrained from acting in accordance with his own beliefs, whether privately or publicly, whether alone or in association with others, within due limits. The council further declares that the right to religious freedom has its foundation in the very dignity of the human person, as this dignity is known through the revealed Word of God and by

reason itself. This right of the human person to religious freedom is to be recognized in the constitutional law whereby society is governed and thus it is to become a civil right. (1965, art. 2)

Catholic doctrine here supports the liberal right to religious freedom for reasons internal to Catholicism. A reasonable Islamic doctrine, and a reasonable atheistic doctrine, might also affirm this same right to religious freedom—not, of course, for the same reasons as Catholic doctrine, but each for its own reasons. In an overlapping consensus, all reasonable comprehensive doctrines will support the right to religious freedom, each for its own reasons. Indeed, in an overlapping consensus, all reasonable comprehensive doctrines will endorse all of a political conception of justice, each from within its own point of view.

Citizens within an overlapping consensus work out for themselves how the liberal “module” fits with their own worldviews. Some citizens may see liberalism as derived directly from their deepest beliefs, as in the quotation from Vatican II above. Others may accept a liberal conception as attractive in itself, but mostly separate from their other concerns. What is crucial is that all citizens view the values of a political conception of justice as very great values, which normally outweigh their other values should these conflict on some particular issue. All citizens, for their own reasons, give the political conception priority in their reasoning about how their society’s basic laws should be ordered.

Rawls sees an overlapping consensus as the most desirable form of stability in a free society. Stability in an overlapping consensus is better than a mere balance of power (a *modus vivendi*) among citizens who hold contending worldviews. After all, power often shifts, and when it does the stability of a *modus vivendi* may be lost.

In an overlapping consensus, citizens affirm a political conception wholeheartedly from within their own perspectives, and so will continue to do so even if their group gains or loses political power. Rawls says that an overlapping consensus is *stable for the right reasons*: each citizen affirms a moral doctrine (a liberal conception of justice) for moral reasons (as given by their comprehensive doctrine). Abiding by liberal basic laws is not a citizen’s second-best option in the face of the power of others; it is each citizen’s first-best option given her own beliefs.

Rawls does not assert that an overlapping consensus is achievable in every liberal society. Nor does he say that, once established, an overlapping consensus must forever endure. Citizens in some societies may have too little in common to converge on a liberal political conception of justice. In other societies, unreasonable doctrines may spread until they overwhelm liberal institutions.

Rawls does hold that history shows both convergence in beliefs and deepening trust among citizens in many liberal societies. This gives hope that an overlapping consensus is at least possible. Where an overlapping consensus is possible, Rawls believes, it is the best support for social stability that a free society can achieve.

## 3.6 Public Reason

Having seen how Rawls answers the challenges of legitimacy and stability, we can return to legitimacy and its criterion of reciprocity: citizens must reasonably believe that all citizens can reasonably accept the enforcement of a particular set of basic laws. It is unreasonable for citizens to attempt to impose what they see as the whole truth on others—political power must be used in ways that all citizens may reasonably be expected to endorse. With his doctrine of *public reason*, Rawls extends this requirement of reciprocity to apply directly to how citizens explain their political decisions to one another. In essence, public reason requires citizens to be able to justify their political decisions to one another using publicly available values and standards.

To take a straightforward example: a Supreme Court justice deciding on a gay marriage law would violate public reason were she to base her opinion on God’s forbidding gay sex in the book of Leviticus, or on a personal spiritual revelation that upholding such a law would hasten the end of days. This is because not all members of society can reasonably be expected to accept Leviticus as stating an authoritative set of political values, nor can a religious premonition be a common standard for evaluating public policy. These values and standards are not public.

Rawls’s doctrine of public reason can be summarized as follows:

*Citizens engaged in certain political activities have a duty of civility to be able to justify their decisions on fundamental political issues by reference only to public values and public standards.*

Each of the highlighted terms in this doctrine can be further elucidated as follows:

•

The **public values** that citizens must be able to appeal to are the values of a political conception of justice: those related to the freedom and equality of citizens, and society as a fair system of cooperation over time. Among such public values are the freedom of religious practice, the political equality of women and of racial minorities, the efficiency of the economy, the preservation of a healthy environment, and the stability of the family (which helps the orderly reproduction of society from one generation to the next). Nonpublic values include values internal to associations like churches (e.g., that women may not hold the highest offices) or private clubs (e.g., that racial minorities can be excluded) which cannot be squared with public values such as these.

•

•

Similarly, citizens should be able to justify their political decisions by **public standards** of inquiry. Public standards are principles of reasoning and rules of evidence that all citizens can reasonably endorse. So citizens should not justify their political decisions by appeal to divination, or to complex and disputed economic or psychological theories. Rather, publicly acceptable standards are those that rely on common sense, on facts generally known, and on the conclusions of science that are well established and not controversial.

•

•

The duty to abide by public reason applies when the most **fundamental political issues** are at stake: issues such as who has the right to vote, which religions are to be tolerated, who will be eligible to own property, and what are suspect classifications for discrimination in hiring decisions. These are what Rawls calls *constitutional essentials and matters of basic justice*. Public reason applies more weakly, if at all, to less momentous political questions, for example to most laws that set rates of taxation, or that put aside public money to maintain national parks.

•

•

Citizens have a duty to constrain their decisions by public reason only when they engage in **certain political activities**, usually when exercising powers of public office. So judges are bound by public reason when they issue their rulings, legislators should abide by public reason when speaking and voting in the legislature, and the executive and candidates for high office should respect public reason in their public pronouncements. Significantly, Rawls says that voters should also heed public reason when they vote. All of these activities are or support exercises of political power, so (by the liberal principle of legitimacy) all must be justifiable in terms that all citizens might reasonably endorse. However, citizens are not bound by any duties of public reason when they engage in other activities, for example when they worship in church, perform on stage, pursue scientific research, send letters to the editor, or talk politics around the dinner table.

•

•

The duty to be able to justify one's political decisions with public reasons is a moral duty, not a legal duty: it is a **duty of civility**. All citizens always have their full legal rights to free expression, and overstepping the bounds of public reason is never in itself a crime. Rather, citizens have a moral duty of mutual respect and civic friendship not to justify their political decisions on fundamental issues by appeal to partisan values or controversial standards of reasoning that cannot be publicly redeemed.

•

In an important proviso, Rawls adds that citizens may speak the language of their controversial comprehensive doctrines—even as public officials, and even on the most fundamental issues—so long as it can be shown that these assertions appeal to public values. So President Lincoln, for instance, could legitimately use Biblical imagery to condemn the evil of slavery, since his condemnations appealed to the public values of freedom and equality. Thus even within its limited range of application, Rawls's doctrine of public reason is rather permissive concerning what citizens may say and do within the bounds of civility.

# 4. Justice as Fairness: Justice within a Liberal Society

Justice as fairness is Rawls's theory of justice for a liberal society. As a member of the family of liberal political conceptions of justice it provides a framework for the legitimate use of political power. Yet legitimacy is only the minimal standard of moral acceptability; a political order can be legitimate without being just. Justice sets the maximal standard: the arrangement of social institutions that is morally best.

Rawls constructs justice as fairness around specific interpretations of the ideas that citizens are free and equal, and that society should be fair. He sees it as resolving the tensions between the ideas of freedom and equality, which have been highlighted both by the socialist critique of liberal democracy and by the conservative critique of the modern welfare state. Rawls also argues that justice as fairness is superior to the dominant tradition in modern political thought: utilitarianism.

In Rawls's egalitarian liberalism, citizens relate to each other as equals within a social order defined by reciprocity, instead of within the unjust status hierarchies familiar from today.

Significant political and economic inequalities are often associated with inequalities of social status that encourage those of lower status to be viewed both by themselves and by others as inferior. This may arouse widespread attitudes of deference and servility, on one side, and a will to dominate and arrogance on the other. These effects of social and economic inequalities can be serious evils and the attitudes they engender great vices... Fixed status ascribed by birth, or by gender or race, is particularly odious (*JF*, 131).

## 4.1 The Basic Structure of Society

Justice as fairness aims to describe a just arrangement of the major political and social institutions of a liberal society: the political constitution, the legal system, the economy, the family, and so on. Rawls calls the arrangement of these institutions a society's *basic structure*. The basic structure is the location of justice because these institutions distribute the main benefits and burdens of social life: who will receive social recognition, who will have which basic rights, who will have opportunities to get what kind of work, what the distribution of income and wealth will be, and so on.

The form of a society's basic structure will have profound effects on the lives of citizens. The basic structure will influence not only citizens' life prospects, but more deeply their goals, their attitudes, their relationships, and their characters. Institutions that will have such pervasive influence on people's lives require justification. Since leaving one's society is not a realistic option for most people, the justification cannot be that citizens have consented to a basic structure by staying in the country. And since the rules of any basic structure will be coercively enforced, often with serious penalties, the demand to justify the imposition of any particular set of rules intensifies further.

In setting out justice as fairness, Rawls assumes that the liberal society in question is marked by reasonable pluralism as described above, and also that it is under reasonably favorable conditions: that there are enough resources for it to be possible for everyone's basic needs to be met. Rawls makes the simplifying assumption that the society is self-sufficient and closed, so that citizens enter it only by birth and leave it only at death. He also confines his attention mainly to ideal theory, putting aside non-ideal theory such as on criminal justice.

## 4.2 Two Guiding Ideas of Justice as Fairness

Social cooperation in some form is necessary for citizens to be able to lead decent lives. Yet citizens are not indifferent to how the benefits and burdens of cooperation will be divided amongst them. Rawls's principles of justice as fairness articulate the central liberal ideas that cooperation should be fair to all citizens regarded as free and as equals. The distinctive interpretation that Rawls gives to these concepts can be seen as combining a negative and a positive thesis.

Rawls's negative thesis starts with the idea that citizens do not deserve to be born into a rich or a poor family, to be born naturally more or less gifted than others, to be born female or male, to be born a member of a particular racial group, and so on. Since these features of persons are morally arbitrary in this sense, citizens are not entitled to more of the benefits of social cooperation simply because of them. For example, the fact that a citizen was born rich, white, and male provides no reason in itself for this citizen to be favored by social institutions.

This negative thesis does not say how social goods should be distributed; it merely clears the decks. Rawls's positive distributive thesis is equality-based reciprocity. All social goods are to be distributed

equally, unless an unequal distribution would be to everyone's advantage. The guiding idea is that since citizens are fundamentally equal, reasoning about justice should begin from a presumption that cooperatively-produced goods should be equally divided. Justice then requires that any inequalities must benefit all citizens, and particularly must benefit those who will have the least. Equality sets the baseline; from there any inequalities must improve everyone's situation, and especially the situation of the worst-off. These strong requirements of equality and reciprocal advantage are hallmarks of Rawls's theory of justice.

## 4.3 The Two Principles of Justice as Fairness

These guiding ideas of justice as fairness are given institutional form by its two principles of justice:

**First Principle:** Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all;

**Second Principle:** Social and economic inequalities are to satisfy two conditions:

1. They are to be attached to offices and positions open to all under conditions of *fair equality of opportunity*;
2. They are to be to the greatest benefit of the least-advantaged members of society (the *difference principle*) (JF, 42-43).

The first principle of equal basic liberties is to be embodied in the political constitution, while the second principle applies primarily to laws governing economic institutions. Fulfillment of the first principle takes priority over fulfillment of the second principle, and within the second principle fair equality of opportunity takes priority over the difference principle.

The first principle affirms that all citizens should have the familiar basic rights and liberties: liberty of conscience and freedom of association, freedom of speech and liberty of the person, the rights to vote, to hold public office, to be treated in accordance with the rule of law, and so on. The first principle accords these rights and liberties to all citizens equally. Unequal rights would not benefit those who would get a lesser share of the rights, so justice requires equal rights for all, in all normal circumstances.

Rawls's first principle confirms widespread convictions about the importance of equal basic rights and liberties. Two further features make this principle distinctive. First is its priority: the basic rights and liberties must not be traded off against other social goods. The first principle disallows, for instance, a policy that would give draft exemptions to college students on the grounds that educated civilians will increase economic productivity. The draft is a drastic infringement on basic liberties, and if a draft is implemented then all who are able to serve must be equally subject to it, even if this means slower growth. Citizens' equal liberty must have priority over economic policy.

The second distinctive feature of Rawls's first principle is that it requires *fair value of the political liberties*. The political liberties are a subset of the basic liberties, concerned with the right to hold public office, the right to affect the outcome of national elections and so on. For these liberties, Rawls requires that citizens should be not only formally but also substantively equal. That is, citizens who are similarly endowed and motivated should have similar opportunities to hold office, to influence elections, and so on regardless of how rich or poor they are. This requirement of the fair value of the political liberties has major implications for how elections should be funded and run, as will be discussed below.

Rawls's second principle of justice has two parts. The first part, fair equality of opportunity, requires that citizens with the same talents and willingness to use them have the same educational and economic opportunities regardless of whether they were born rich or poor. "In all parts of society there are to be roughly the same prospects of culture and achievement for those similarly motivated and endowed" (JF, p. 44).

So, for example, if we assume that natural endowments and the willingness to use them are evenly distributed across children born into different social classes, then within any type of occupation (generally specified) we should find that roughly one quarter of people in that occupation were born into the top 25% of the income distribution, one quarter were born into the second-highest 25% of the income distribution, one quarter were born into the second-lowest 25%, and one-quarter were born into the lowest 25%. Since class of origin is a morally arbitrary fact about citizens, justice does not allow class of origin to turn into unequal opportunities for education or meaningful work.

The second part of the second principle is the difference principle, which regulates the distribution of wealth and income. Allowing inequalities of wealth and income can lead to a larger social product:

higher wages can cover the costs of training and education, for example, and can provide incentives to fill jobs that are more in demand. The difference principle allows inequalities of wealth and income, so long as these will be to everyone's advantage, and specifically to the advantage of those who will be worst off. The difference principle requires, that is, that any economic inequalities be to the greatest advantage of those who are advantaged least.

To illustrate, consider four hypothetical economic structures A-D, and the lifetime-average levels of income that these different economic structures would yield for representative members of three groups:

<i>Economy</i>	<i>Least-Advantaged Group</i>	<i>Middle Group</i>	<i>Most-Advantaged Group</i>
A	10,000	10,000	10,000
B	12,000	30,000	80,000
C	30,000	90,000	150,000
D	20,000	100,000	500,000

Here the difference principle selects Economy C, because it contains the distribution where the least-advantaged group does best. Inequalities in C are to everyone's advantage relative to a completely equal distribution (Economy A), and relative to a more equal distribution (Economy B). But the difference principle does not allow the rich to get richer at the expense of the poor (Economy D). The difference principle embodies equality-based reciprocity: from an egalitarian baseline, it requires that any inequalities are good for all, and especially for the worst-off.

The difference principle is partly based on the negative thesis that the distribution of natural assets is undeserved. A citizen does not merit more of the social product simply because she was lucky enough to be born with the potential to develop skills that are currently in high demand. Yet this does not mean that everyone must get the same shares. The fact that citizens have different talents and abilities can be used to make everyone better off. In a society governed by the difference principle, citizens regard the distribution of natural endowments as a common asset that can benefit all. Those better endowed are welcome to use their gifts to make themselves better off, so long as their doing so also contributes to the good of those less well endowed.

The difference principle thus expresses a positive ideal, an ideal of social unity. In a society that satisfies the difference principle, citizens know that their economy works to everyone's benefit, and that those who were lucky enough to be born with greater natural potential are not getting richer at the expense of those who were less fortunate. One might contrast Rawls's positive ideal to Nozick's ideal of libertarian freedom, or to meritocratic ideas about economic justice that are dominant within many democracies. "In justice as fairness," Rawls says, "men agree to share one another's fate" (TJ, 102).

## 4.4 The Conception of Citizens

Having surveyed Rawls's two principles of justice as fairness, we can return to Rawls's interpretations of the liberal ideas that citizens are free and equal, and that society should be fair. Rawls uses these conceptions of citizens and society to construct the official justification for the two principles: the argument from the original position.

Rawls's interpretation of the idea that citizens are free is as follows. Citizens are free in that each sees herself as being entitled to make claims on social institutions in her own right—citizens are not slaves or serfs, dependent for their social status on others. Citizens are also free in that they see their public identities as independent of any particular comprehensive doctrine: a citizen who converts to Islam, or who recants her faith, will expect, for example, to retain all her political rights and liberties throughout the transition. Finally, citizens are free in being able to take responsibility for planning their own lives, given the opportunities and resources that they can reasonably expect.

Citizens are equal, Rawls says, in virtue of having the capacities to participate in social cooperation over a complete life. Citizens may have greater or lesser skills, talents, and powers "above the line" that cooperation requires, but differences above this line have no bearing on citizens' equal political status.

Rawlsian citizens are not only free and equal, they are also reasonable and rational. The idea that citizens are reasonable is familiar from political liberalism. Reasonable citizens have the capacity to abide by fair terms of cooperation, even at the expense of their own interests, provided that others are also willing to do so. In justice as fairness, Rawls calls this reasonableness the capacity for a *sense of justice*. Citizens are also rational: they have the capacity to pursue and revise their own view of what is valuable in human life. Rawls calls this the capacity for a *conception of the good*. Together these capacities are called the *two moral powers*.

Like every theory of justice (for example those of Locke, Rousseau, and Mill), justice as fairness requires an account of citizens' fundamental interests: what citizens need *qua* citizens. Rawls derives his account of *primary goods* from the conception of the citizen as free and equal, reasonable and rational. Primary goods are essential for developing and exercising the two moral powers, and are useful for pursuing a wide range of specific conceptions of the good life. Primary goods are these:

- The basic rights and liberties;
- Freedom of movement, and free choice among a wide range of occupations;
- The powers of offices and positions of responsibility;
- Income and wealth;
- The social bases of self-respect: the recognition by social institutions that gives citizens a sense of self-worth and the confidence to carry out their plans (*JF*, 58–59).

All citizens are assumed to have fundamental interests in getting more of these primary goods, and political institutions are to evaluate how well citizens are doing according to what primary goods they have. It is equalities and inequalities of these primary goods that, Rawls claims, are of the greatest political significance.

## 4.5 The Conception of Society

Rawls's conception of society is defined by fairness: social institutions are to be fair to all cooperating members of society, regardless of their race, gender, religion, class of origin, natural talents, reasonable conception of the good life, and so on.

Rawls also emphasizes *publicity* as an aspect of fairness. In what he calls a *well-ordered society* all citizens accept the principles of justice and know that their fellow citizens also do so, and all citizens recognize that the basic structure is just. The full philosophical justifications for the principles of justice are also knowable by and acceptable to all reasonable citizens.

The idea behind publicity is that since the principles for the basic structure will be coercively enforced on free citizens, they should stand up to public scrutiny. The publicity condition requires that a society's operative principles of justice not be too esoteric, and not be screens for deeper power relations. Fairness requires that, in "public political life, nothing need be hidden... there is no need for the illusions and delusions of ideology for society to work properly and for citizens to accept it willingly" (*PL*, 68–69).

## 4.6 The Original Position

Rawls's conceptions of citizens and society are still quite abstract, and some might think innocuous. The *original position* aims to move from these abstract conceptions to determinate principles of social justice. It does so by translating the question: "What are fair terms of social cooperation for free and equal citizens?" into the question "What terms of cooperation would free and equal citizens agree to under fair conditions?" The move to agreement among citizens is what places Rawls's justice as fairness within the social contract tradition of Locke, Rousseau, and Kant.

The strategy of the original position is to construct a method of reasoning that models abstract ideas about justice so as to focus their power together onto the choice of principles. So Rawls's conceptions of citizens and of society are built into the design of the original position itself. Rawls's intent is that readers will see the outcome of the original position as justified because they will see how it embodies plausible understandings of citizens and society, and also because this outcome confirms many of their considered convictions about justice on specific issues.

The original position is a thought experiment: an imaginary situation in which each real citizen has a representative, and all of these representatives come to an agreement on which principles of justice should order the political institutions of the real citizens. This thought experiment is better than trying to get all real citizens actually to assemble in person to try to agree to principles of justice for their society. Even if that were possible, the bargaining among real citizens would be influenced by all sorts of factors irrelevant to justice, such as who could threaten the others most, or who could hold out for longest.

The original position abstracts from all such irrelevant factors. The original position is a fair situation in which each citizen is represented as only a free and equal citizen: each representative wants only what free and equal citizens want, and each tries to agree to principles for the basic structure while situated fairly with respect to the other representatives. The design of the original position thus models the ideas of freedom, equality, and fairness. For example, fairness and equality are modeled in the original position by the symmetrical situation of the parties: no citizen's representative is able to threaten any other citizen's representative, or to hold out longer for a better deal.

The most striking feature of the original position is the *veil of ignorance*, which prevents arbitrary facts about citizens from influencing the agreement among their representatives. As we have seen, Rawls holds that the fact that a citizen is of a certain race, class, and gender is no reason for social institutions to favor or disfavor her. Each representative in the original position is therefore deprived of knowledge of the race, class, and gender of the real citizen that they represent. In fact, the veil of ignorance deprives the parties of all facts about citizens that are irrelevant to the choice of principles of justice: not only facts about their race, class, and gender but also facts about their age, natural endowments, and more. Moreover the veil of ignorance also screens out specific information about what society is like right now, so as to get a clearer view of the permanent features of a just social system.

Behind the veil of ignorance, the informational situation of the parties that represent real citizens is as follows:

- Parties do not know

- The race, ethnicity, gender, age, income, wealth, natural endowments, comprehensive doctrine, etc. of any of the citizens in society, or to which generation in the history of the society these citizens belong;
- The political system of the society, its class structure, economic system, or level of economic development.
- Parties do know
  - That citizens in the society have different comprehensive doctrines and plans of life; that all citizens have interests in more primary goods;
  - That the society is under conditions of moderate scarcity: there is enough to go around, but not enough for everyone to get what they want;
  - General facts and common sense about human social life; general conclusions of science (including economics and psychology) that are uncontroversial.

The veil of ignorance situates the representatives of free and equal citizens fairly with respect to one another. No party can press for agreement on principles that will arbitrarily favor the particular citizen they represent, because no party knows the specific attributes of the citizen they represent. The situation of the parties thus embodies reasonable conditions, within which the parties can make a rational agreement. Each party tries to agree to principles that will be best for the citizen they represent (i.e., that will maximize that citizen's share of primary goods). Since the parties are fairly situated, the agreement they reach will be fair to all actual citizens.

The design of the original position also models other aspects of Rawls's conceptions of citizens and society. For example, the publicity of a well-ordered society is modeled by the fact that the parties must choose among principles that can be publicly endorsed by all citizens. There are also some assumptions that make the hypothetical agreement determinate and decisive: the parties are not motivated by envy (i.e., by how much citizens besides their own end up with); the parties are neither risk-seeking nor risk-averse; and the parties must make a final agreement on principles for the basic structure: there are no "do-overs" after the veil of ignorance is lifted and the parties learn which real citizen they represent.

## 4.7 The Argument from the Original Position: The Selection of Principles

The argument from the original position has two parts. In the first part, the parties agree to principles of justice. In the second part, the parties check that a society ordered by these principles could be stable over time. Rawls only attempts to show that his two principles of justice as fairness would be favored over utilitarian principles, since he sees utilitarianism as the main competing tradition of reasoning about justice. The parties are thus presented with a choice between Rawls's two principles and utilitarian principles, and asked which principles they would prefer to agree to.

The first part of the original position contains two fundamental comparisons between Rawls's principles and utilitarian principles. In the first comparison, the parties compare Rawls's principles to *the principle of average utility*: the principle that the basic structure should be arranged so as to produce the highest level of utility averaged among all citizens. Rawls argues that the parties would favor his principles in this comparison, because the first principle of justice as fairness secures equal liberties for all citizens.

In this first comparison, Rawls argues that it is rational for the parties to use *maximin* reasoning: to maximize the minimum level of primary goods that the citizens they represent might find themselves with. And maximin reasoning, he says, favors justice as fairness.

Under average utilitarianism, Rawls argues, the basic liberties of some citizens might be restricted for the sake of greater benefits to other citizens. For example, restricting the political and religious liberties of a weak minority might work to the benefit of the majority, and so produce a higher average level of utility in the society. A party in the original position will find the possibility that their citizen might be denied political and religious liberties intolerable, given that the party could instead secure equal liberties for their citizen by choosing justice as fairness. A party will not be willing to gamble with the political standing and deepest commitments of the citizen they represent, Rawls says, when they could safeguard the standing and commitments of their citizen even if their citizen turns out to be in a weak minority.

Moreover, Rawls says, a society ordered by the principles of justice as fairness has other advantages over a utilitarian society. Securing equal basic liberties for all encourages a spirit of cooperation among citizens on a basis of mutual respect, and takes divisive conflicts about whether to deny liberties to some citizens off the political agenda. By contrast, a utilitarian society would be riven by mutual suspicions, as different groups put forward highly speculative arguments that average utility could be increased by implementing their partisan policies. Rawls's first principle, by securing permanent equal liberties for all citizens, increases social harmony by making it easier for justice to be seen to be done. The balance of considerations in favor of justice as fairness over average utility here is, Rawls claims, decisive.

In the second fundamental comparison, the parties are offered a choice between justice as fairness and *the principle of restricted utility*. The principle of restricted utility is identical to Rawls's two

principles, except that the difference principle is replaced with a principle which says that the distribution of wealth and income should maximize average utility, constrained by a guaranteed minimum level of income for all. While the first comparison turned on the importance of the basic liberties, the second comparison contains Rawls's formal argument for the difference principle.

Maximin reasoning plays no role in the argument for the difference principle. Nor does aversion to uncertainty (*JF*, xvi, 43, 95, 96).

In this second comparison, Rawls argues that the parties will favor justice as fairness because its principles provide a better basis for enduring cooperation among all citizens. The difference principle, he says, asks less of the better-off than restricted utility asks of the worst-off. Under the difference principle, he says, those who are better endowed are permitted to gain more wealth and income, on the condition that their doing so also benefits their fellow citizens. Under restricted utility, by contrast, those living at the minimum will suspect that their interests have been sacrificed to make the better-off better off still. These citizens at the minimum may become cynical about their society, and withdraw from active participation in public life.

Moreover, it is again difficult to maintain a public agreement as to which economic policies actually will maximize average utility, and debates over where to set the guaranteed minimum may lead to mistrust among social classes. The difference principle instead encourages mutual trust and the cooperative virtues by instantiating an ideal of economic reciprocity. Each party will see the advantages for the citizen they represent of securing the more harmonious social world of justice as fairness.

## 4.8 The Argument from the Original Position: The Check for Stability

Having selected the two principles of justice as fairness, the parties turn to the second part of the original position: the check that these principles can order a society stably over time. The parties check, that is, whether those who grow up under institutions arranged by these principles will develop sufficient willingness to abide by them that the principles can serve as the focus of an enduring overlapping consensus.

Rawls argues that the parties will see that his two principles are congruent with each citizen's good. Under the two principles, the society's basic institutions affirm the freedom and equality of each citizen, giving a public basis for each citizen's self-respect. This public basis of self-respect is vital for citizens to be able to pursue their life plans with energy and confidence. Citizens will also see that the basic liberties allow sufficient social space for them to pursue their reasonable conceptions of the good. Whether poor or rich, citizens will tend not to be envious or imperious, as they will see how the economy works toward the reciprocal advantage of all. And citizens may be satisfied by reflecting on the collective good that they can achieve with each other, by working together to maintain just institutions over time.

Given that the two principles are congruent with citizens' good, Rawls argues that it is reasonable to suppose that citizens will develop a desire to act in accordance with them. People become attached to people and institutions that they see benefiting them, and the two principles create a social world in which each citizen can pursue her own ends on a basis of mutual respect with other citizens. Since this is experienced as a good, the principles will gain citizens' willing and stable allegiance. "The most stable conception of justice," Rawls says, "is one that is perspicuous to our reason, congruent with our good, and rooted not in abnegation but in affirmation of the self" (*TJ*, 261).

## 4.9 Just Institutions: The Four-Stage Sequence

The two parts of the argument for justice as fairness above occur at the first stage of the original position. At this first stage, the parties also agree to a *principle of just savings* to regulate how much each generation must save for future generations. Since the parties do not know which era the citizens they represent live in, it is rational for them to choose a savings principle that is fair to all generations. Rawls says that the parties need not choose a savings principle that requires endless economic growth. Rather, the parties may prefer a Millian "steady state" of zero real growth, once a generation has been reached in which the two principles are satisfied.

After agreeing on the two principles and a principle of just savings, the parties then proceed further through the *four-stage sequence*, tailoring these general principles to the particular conditions of the society of the citizens they represent. Through this four-stage sequence, the veil of ignorance that screens out information about society's general features gradually becomes thinner, and the parties use the new information to decide on progressively more determinate applications of the principles already agreed upon. The parties, that is, progressively fill in the institutional details of what justice requires in the real world.

At the second stage of the original position, the parties are given more information about society's political culture and economic development, and take on the task of crafting a *constitution* that realizes the two principles of justice. At the third stage, the parties learn still more about the details of society, and agree to specific *legislation* that realize the two principles within the constitutional framework decided at the second stage. At the fourth stage, the parties have full information about

society, and reason as *judges and administrators* to apply the previously-agreed legislation to particular cases. When the four stages are complete, the principles of justice as fairness are fully articulated for society's political life.

To illustrate: at the constitutional (second) and legislative (third) stages, the parties specify basic liberties such as "freedom of thought" into more particular rights, like the right to free political speech. The right to political speech is itself then further specified as the right to criticize the government, the rights protecting the press from political interference, and so on. Through the four-stage sequence, the parties also adjust the basic liberties to fit with one another and with other values, always aiming for an overall scheme of liberties that will best enable citizens to develop and exercise their two moral powers and pursue their determinate conceptions of the good (*PL*, 289–371).

At the later stages, the parties also work out the institutions that will be necessary to realize the fair value of the equal political liberties. On this topic, Rawls is adamant: unless there are public funds for elections, restrictions on campaign contributions, and substantially equal access to the media, politics will be captured by concentrations of private economic power. This will make it impossible for equally-able citizens to have equal opportunities to influence politics regardless of their wealth, as fair value demands. As he says in a discussion of public reason, public deliberation must be "set free from the curse of money" (*PL*, 449).

The parties attempt to realize the second principle of justice at the legislative stage, by shaping the laws that regulate property, contract, taxation, inheritance, hiring, minimum wages, and so on. Their task is not to allocate some fixed set of goods that appear from nowhere, but rather to devise a set of institutions for education, production, and distribution whose operation will realize fair equality of opportunity and the difference principle over time.

For fair equality of opportunity, Rawls emphasizes that laws and policies must go beyond merely preventing discrimination in education and hiring. To ensure fair opportunity regardless of social class of origin, the state must also fund high-quality education for the less well-off. Moreover, the state must also guarantee both a basic minimum income and health care for all.

On realizing the difference principle, Rawls says that the goal is an economic order that maximizes the position of the worst-off group (e.g., unskilled laborers, or those with less than half of the median wealth and income over their lifetimes). Given that institutions realizing the prior principles are already in place, this should be approximately achievable by, for example, varying marginal rates of tax and tax exemptions.

Rawls explicitly rejects the welfare state (*JF*, 137–40). Welfare-state capitalism leaves control of the economy in the hands of a group of rich private actors. It therefore fails to ensure for all citizens enough resources to have roughly equal chances of influencing politics, or to have sufficiently equal opportunities in education and employment. The welfare state therefore tends to generate a demoralized under-class.

Laissez-faire capitalism is even worse for equality than the welfare state along these dimensions. And a socialist command economy would put too much power in the hands of the state, again endangering political equality and also threatening basic liberties such as free choice of employment.

Justice as fairness, Rawls says, favors either a *property-owning democracy* or liberal (democratic) socialism. The government of a property-owning democracy takes steps to encourage widespread ownership of productive assets and broad access to education and training. Liberal socialism is similar, but features worker-managed firms. The aim of both systems of political economy is to enable all citizens, even the least advantaged, to manage their own affairs within a context of significant social and economic equality. "The least advantaged are not, if all goes well, the unfortunate and unlucky—objects of our charity and compassion, much less our pity—but those to whom reciprocity is owed as a matter of basic justice" (*JF*, 139).

## 4.10 The Original Position and Political Constructivism

Rawls describes the original position as a useful device for reaching greater reflective equilibrium. He holds that the value of the original position as a method of reasoning is affirmed when it selects the first principle of justice, since the first principle accords with many people's settled convictions about the importance of assuring the basic rights and liberties for all. Having gained credibility by confirming these settled moral judgments, the original position then goes on to select principles for issues on which people's judgments may be less certain, such as how society should structure employment opportunities, and what a just distribution of wealth and income might be.

In this way, the original position first confirms and then extends common judgments about justice. For Rawls it is important that the same method of reasoning that explains the equal basic liberties also justifies more political and economic equality than many people might have initially expected. The momentum of the argument for the first principle carries through to the argument for the second principle. Those who believe in equal basic liberties, but who reject the other egalitarian features of justice as fairness, must try to find some other route to justifying those basic liberties.

The original position is also the crux of Rawls's metaethical theory, *political constructivism*. Political constructivism is Rawls's account of the objectivity and validity of political judgments.

The original position embodies, Rawls says, all of the relevant conceptions of person and society, and principles of practical reasoning, for making judgments about justice. When there is an overlapping consensus focused on justice as fairness, the original position specifies a shared public perspective from which all citizens can reason about the principles of justice and their application to their society's institutions. Judgments made from this perspective are then objectively correct, in the sense of giving reasons to citizens to act regardless of their actual motivations, or the reasons they think they have within their particular points of view.

Political constructivism does not maintain that the original position shows that the principles of justice as fairness are true. Questions of truth are ones about which reasonable citizens may disagree, and are to be addressed by each citizen from within her own comprehensive doctrine. Judgments made from the original position are, however, valid, or as Rawls says, reasonable.

## 5. The Law of Peoples: Liberal Foreign Policy

With the theories of legitimacy and justice for a self-contained liberal society completed, Rawls then extends his approach to international relations with the next in his sequence of theories: the law of peoples.

Rawls assumes that no tolerable world state could be stable. He cites Kant in asserting that a world government would either be a global despotism or beleaguered by groups fighting to gain their political independence. So the law of peoples will be international, not cosmopolitan: it will be a foreign policy that guides a liberal society in its interactions with other societies, both liberal and non-liberal.

Rawls describes the main ideas motivating his law of peoples as follows:

Two main ideas motivate the Law of Peoples. One is that the great evils of human history—unjust war and oppression, religious persecution and the denial of liberty of conscience, starvation and poverty, not to mention genocide and mass murder—follow from political injustice, with its own cruelties and callousness... The other main idea, obviously connected with the first, is that, once the gravest forms of political injustice are eliminated by following just (or at least decent) social policies and establishing just (or at least decent) basic institutions, these great evils will eventually disappear (*LP*, 6-7).

The most important feature of the “realistic utopia” that Rawls envisages in *The Law of Peoples* is that the great evils of human history no longer occur. The most important condition for this realistic utopia to come about is that all societies are internally well-ordered: that all have just, or at least decent, domestic political institutions.

### 5.1 The International Basic Structure and the Principles of the Law of Peoples

Much of Rawls's presentation of the law of peoples parallels the presentations of political liberalism and justice as fairness. As a liberal society has a basic structure of institutions so, Rawls says, there is an international basic structure (*LP*, 33, 62, 114, 115, 122, 123). While Rawls does not say that the international basic structure has a pervasive impact on the life chances of individuals, the rules of this basic structure are coercively enforced (for example, Iraq's invasion of Kuwait in 1990 was coercively reversed by a coalition of other countries). The principles that should regulate this international basic structure thus require justification. The justification of these principles must accommodate the fact that there is even more pluralism in worldviews among contemporary societies than there is within a single liberal society.

Rawls puts forward eight principles for ordering the international basic structure:

1. Peoples are free and independent, and their freedom and independence are to be respected by other peoples.
2. Peoples are to observe treaties and undertakings.

3. Peoples are equal and are parties to the agreements that bind them.
4. Peoples are to observe the duty of nonintervention (except to address grave violations of human rights).
5. Peoples have a right of self-defense, but no right to instigate war for reasons other than self-defense.
6. Peoples are to honor human rights.
7. Peoples are to observe certain specified restrictions in the conduct of war.
8. Peoples have a duty to assist other peoples living under unfavorable conditions that prevent their having a just or decent political and social regime (*LF*, 37).

All of these principles, with the exception of the last one, are familiar from contemporary international law (though Rawls's list of human rights for principles 4 and 6 is shorter than the list in international law). Rawls also leaves room for his law of peoples to accommodate various organizations that may help societies to increase their political and economic coordination, such as idealized versions of a United Nations, a World Trade Organization, and a World Bank.

## 5.2 Peoples: Liberal and Decent

The actors in Rawls's international theory are not individuals (citizens) but societies (*peoples*). A people is a group of individuals ruled by a common government, bound together by common sympathies, and firmly attached to a common conception of right and justice. "People" is a moralized concept, and not all states currently on the world map qualify as such.

Rawls's conception of peoples within the law of peoples parallels his conception of citizens within justice as fairness. Peoples see themselves as free in the sense of being rightfully politically independent; and as equal in regarding themselves as equally deserving of recognition and respect. Peoples are reasonable in that they will honor fair terms of cooperation with other peoples, even at cost to their own interests, given that other peoples will also honor those terms. Reasonable peoples are thus unwilling to try to impose their political or social ideals on other reasonable peoples. They satisfy the criterion of reciprocity with respect to one another.

Rawls describes the fundamental interests of a people as follows:

- Protecting its political independence, its territory, and the security of its citizens;
- Maintaining its political and social institutions and its civic culture;
- Securing its proper self-respect as a people, which rests on its citizens' awareness of its history and cultural accomplishments.

Rawls contrasts peoples with *states*. A state, Rawls says, is moved by the desires to enlarge its territory, or to convert other societies to its religion, or to enjoy the power of ruling over others, or to increase its relative economic strength. Peoples are not states, and as we will see peoples may treat societies that act like states as international outlaws.

Peoples are of two types, depending on the nature of their domestic political institutions. *Liberal peoples* satisfy the requirements of political liberalism: they have legitimate liberal constitutions, and they have governments that are under popular control and not driven by large concentration of private economic power.

*Decent peoples* are not internally just from a liberal perspective. Their basic institutions do not recognize reasonable pluralism or embody any interpretation of the liberal ideas of free and equal citizens cooperating fairly. The institutions of a decent society may be organized around a single comprehensive doctrine, such as a dominant religion. The political system may not be democratic, and women or members of minority religions may be excluded from public office. Nevertheless, decent peoples are well-ordered enough, Rawls says, to merit equal membership in international society.

Like all peoples, decent peoples do not have aggressive foreign policies. Beyond this, Rawls describes one type of decent society—a *decent hierarchical society*—to illustrate what decency requires.

A decent hierarchical society's basic structure specifies a decent system of social cooperation. First, it secures a core list of *human rights*. Second, its political system takes the fundamental interests of all persons into account through a *decent consultation hierarchy*. This means that the government genuinely consults with the representatives of all social groups, which together represent all persons in the society, and that the government justifies its laws and policies to these groups. The government does not close down protests, and responds to any protests with conscientious replies. The government also supports the right of citizens to emigrate.

Rawls imagines a decent hierarchical society that he calls "Kazanistan." In Kazanistan, Islam is the favored religion, and only Muslims can hold the high office. However non-Muslim religions may be practiced without fear, and believers in them are encouraged to take part in the civic culture of the wider society. Minorities are not subject to arbitrary discrimination by law, or treated as inferior by Muslims. Kazanistan would qualify, Rawls says, as a decent, well-ordered member of the society of peoples, entitled to respectful toleration and equal treatment by other peoples.

## 5.3 International Toleration and Human Rights

Liberal peoples tolerate decent peoples, and indeed treat them as equals. Not to do so, Rawls says, would be to fail to express sufficient respect for acceptable ways of ordering a society. Liberal peoples should recognize the good of national self-determination, and let decent societies decide their futures for themselves. The government of a liberal people should not criticize decent peoples for failing to be liberal, or set up incentives for them to become more so. Criticism and inducements may cause bitterness and resentment within decent peoples, and so be counter-productive.

Indeed public reason imposes duties of civility upon the members of international society, just as it does upon members of a liberal society. Government officials and candidates for high office should explain their foreign policy positions to other peoples in terms of the principles and values of the law of peoples, and should avoid reliance on contentious parochial reasons that all peoples cannot reasonably share.

One major reason that liberal peoples tolerate decent peoples, Rawls says, is that decent peoples secure for all persons within their territory a core list of human rights. These core human rights include rights to subsistence, security, personal property, and formal equality before the law, as well as freedoms from slavery, protections of ethnic groups against genocide, and some measure of liberty of conscience (but not, as we have seen, a right to democratic participation). These core human rights are the minimal conditions required for persons to be able to engage in social cooperation in any real sense, so any well-ordered society must protect them.

The role of human rights in the law of peoples is thus to set limits on international toleration. Any society that guarantees Rawls's list of human rights is to be immune from coercive intervention from other peoples. Societies that violate human rights overstep the limits of toleration, and may rightly be subject to economic sanctions or even military intervention.

## 5.4 The International Original Position

The international original position parallels the domestic original position of justice as fairness. This original position answers the question: "What terms of cooperation would free and equal peoples (liberal and decent) agree to under fair conditions?" The strategy is to build the conception of peoples into the design of this original position, along with restrictions on reasons for favoring basic principles of international law. The strategy, that is, is to describe reasonable conditions under which a rational agreement on principles can be made.

In the international original position, representatives of each people agree on principles for the international basic structure. Each party is behind a veil of ignorance, deprived of information about the people they represent, such as the size of its territory and population, and its relative political and economic strength. Each party tries to do the best they can for the people they represent, in terms of the fundamental interests that all peoples have.

Rawls claims that the parties in the international original position would favor the eight principles listed above. Starting from a baseline of equality and independence, the parties would see no reason to introduce inequalities into the relationships among peoples (beyond certain functional inequalities in the design of cooperative organizations, such as richer countries contributing more to an idealized United Nations). The parties would reject international utilitarian principles, as no people is prepared to accept that it should sacrifice its fundamental interests for the sake of greater total global utility.

After selecting the eight principles of the law of peoples, the parties next check that these principles can stably order international relations over time. Analogously to the domestic case, the parties will see that the principles of the law of peoples affirm the good of peoples, and that peoples will develop trust and confidence in one another as all willingly continue to abide by these principles. The stability of the international political order will thus be stability for the right reasons (and not a mere *modus vivendi*), since each people will affirm the principles as its first-best option whatever the international balance of power might become.

Rawls also attempts to draw empirical support for his stability argument from the literature on the democratic peace. Social scientists have found that historically democracies have tended not to go to war with one another. Rawls explains this by saying that liberal societies are, because of their internal political structures, *satisfied*. Liberal peoples have no desires for imperial glory, territorial expansion, or to convert others to their religion, and whatever goods and services they need from other countries they can obtain through trade. Liberal peoples, Rawls says, have no reasons to fight aggressive wars, so a genuine peace can endure among them. And since decent peoples are defined as non-aggressive, any decent people can join this liberal peace as well.

Once the parties have agreed to the eight principles of the law of peoples, they then continue to specify these principles more precisely in a process analogous to the domestic four-stage sequence.

## 5.5 Non-Ideal Theory: Outlaw States and Burdened Societies

The principles selected in the international original position contain provisions for non-ideal situations: situations in which nations are unwilling to comply with the ideal principles, or are unable to cooperate on their terms. These provisions are embedded in principles 4 through 8 of the law of peoples.

*Outlaw states* are non-compliant: they threaten the peace by attempting to expand their power and influence, or by violating the human rights of those within their territory. The principles of the law of peoples allow peoples to fight these outlaw states in self-defense, and to take coercive actions against them to stop their violations of human rights. In any military confrontations with outlaw states, peoples must obey the principles of the just prosecution of war, such as avoiding direct attacks on enemy civilians in all but the most desperate circumstances. The aim of war, Rawls says, must be to bring all societies to honor the law of peoples, and eventually to become fully participating members of international society.

*Burdened societies* struggle with social and economic conditions that make it difficult for them to maintain either liberal or decent institutions. A burdened society may lack sufficient material or social resources to support a scheme of social cooperation, perhaps because its population has grown beyond its territory's means to support it. It is the basic structure and political culture of a society that are most crucial for its self-sufficiency; the international community must help a burdened society to rise above this threshold. The law of peoples (eighth principle) requires that burdened peoples be assisted until they can handle their own affairs (i.e., become well-ordered).

This duty of assistance is Rawls's greatest divergence from the rules of today's international law. Accepting this duty would require significant changes in how nations respond to global poverty and failed states.

## 5.6 Reconciliation and Realistic Utopia

Rawls's vision is of a perpetually peaceful and cooperative international order, where liberal and decent peoples stand ready to pacify aggressive states, to secure core human rights, and to help struggling countries until they become self-sufficient.

Compared to other theories, Rawls's international theory has limited ambitions. Officials of democratic societies can do little more than hope that decent societies will become internally more tolerant and democratic. Once the duty to assist burdened peoples is satisfied, there are no further requirements on international economic distributions: for Rawls, international economic inequalities are of no political concern as such. Moreover, individuals around the world may suffer greatly from bad luck, and they may be haunted by spiritual emptiness. These are not concerns reached by a Rawlsian foreign policy.

The limited practical goal of Rawls's law of peoples is the elimination of the great evils of human history: unjust war and oppression, religious persecution and the denial of liberty of conscience, starvation and poverty, genocide and mass murder. The limits of this ambition mean that there will be much in the world to which Rawls's political philosophy offers no reconciliation.

Nevertheless, while Rawls's vision is limited, it is also utopian. To believe that Rawls's vision is possible is to believe that individuals are not inevitably selfish or amoral, and that international relations can be more than merely a contest for domination, wealth, and glory. Human life, and human history, need not be merely a Hobbesian "perpetual and restless desire of power after power" (1651, 58). Affirming the possibility of a just and peaceful future can inoculate us against a cynicism that undermines the decency, reciprocity, and reasonableness that exist now and that may grow from now on.

"By showing how the social world may realize the features of a realistic utopia," Rawls says, "political philosophy provides a long-term goal of political endeavor, and in working toward it gives meaning to what we can do today" (LF, 128).